|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Методы анализа данных»

**Практическая работа № 5**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-05-22, Новиков Денис Владиславович* | (подпись) | |
| Преподаватель | *Шкерин Алексей Леонидович* | (подпись) | |
| Отчет представлен | «11» сентября 2024 г. | |  | |

Москва 2024 г.

## Постановка задачи:

Использовать сервисы Kafka в командной строке для создания темы. Также использовать Kafka для создания производителей и потребителей и передачи данных через них.

## Ход работы:

### Создание темы, производителя и потребителя Kafka

Для того, чтобы сократить нагрузку на нашу виртуальную машину, мы остановим HBase и запустим только Apache Kafka. Остановим сервисы HBase. Перезапустим Kafka и Zookeeper на Рисунке 1.

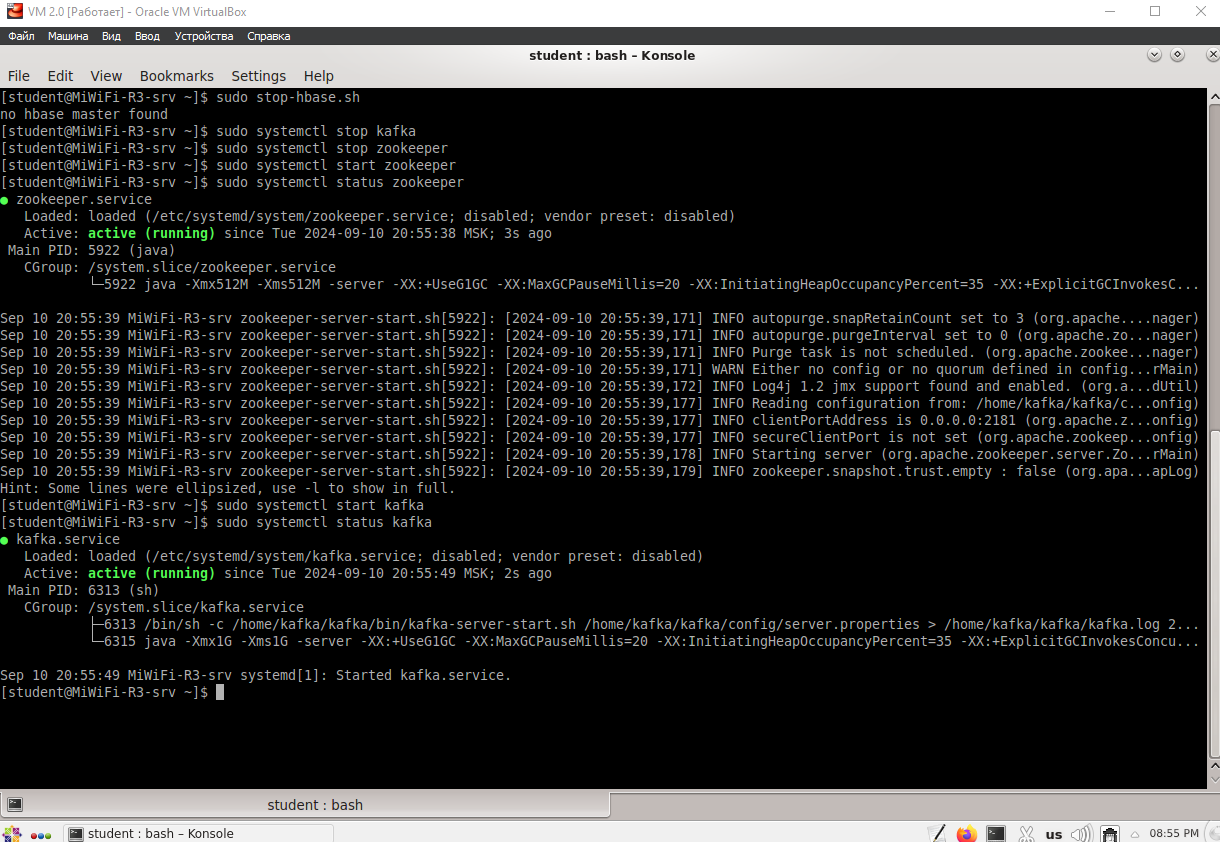


Рисунок 1 — Настройка Apache Kafka

Создадим тему topic1\_logs в терминале на Рисунке 2.

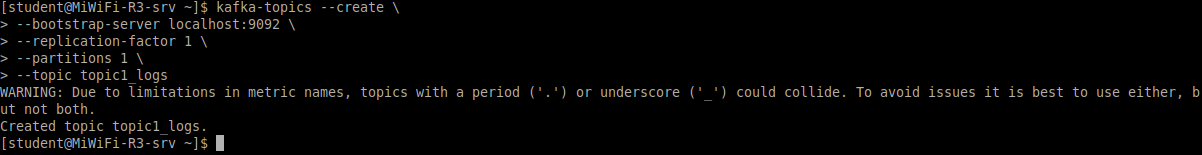


Рисунок 2 — Создание темы topic1\_logs в терминале

Для отображения всех тем kafka, используем ключ --list, и убедимся, что тема, которую мы только что создали, отображена в списке на Рисунке 3.

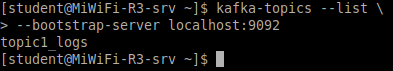


Рисунок 3 — Отображение всех тем kafka

Просмотрим тему topic1\_logs на Рисунке 4.

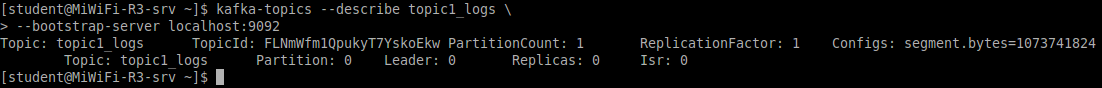


Рисунок 4 — Тема topic1\_logs

Откроем окно терминала и создадим производителя в одном окне, а потребителя в другом. Из первого терминала запустим команду kafka-console-producer для запуска производителя на Рисунке 5.

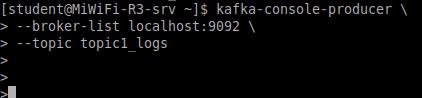


Рисунок 5 — Запуск производителя

Из второго терминала вызовем команду kafka-console-consumer для создания потребителя на Рисунке 6.

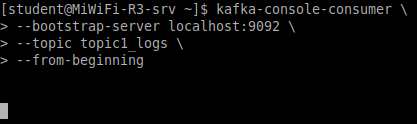


Рисунок 6 — Запуск потребителя

В терминале производителя выберем Edit -> Rename Tab и введем имя «Producer», аналогично для второго терминала «Consumer» на Рисунке 7.

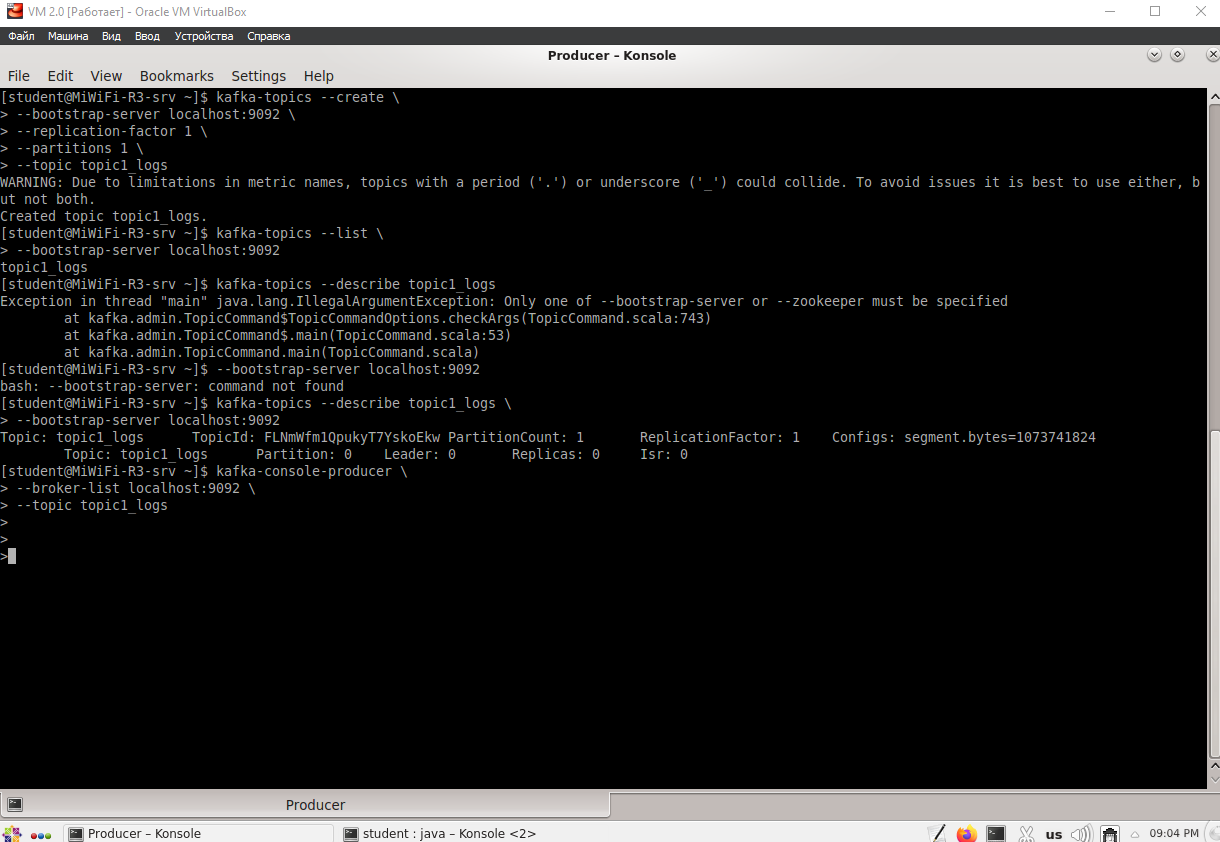


Рисунок 7 — Подготовленный терминал Producer

Подготовленный терминал Consumer на Рисунке 8.

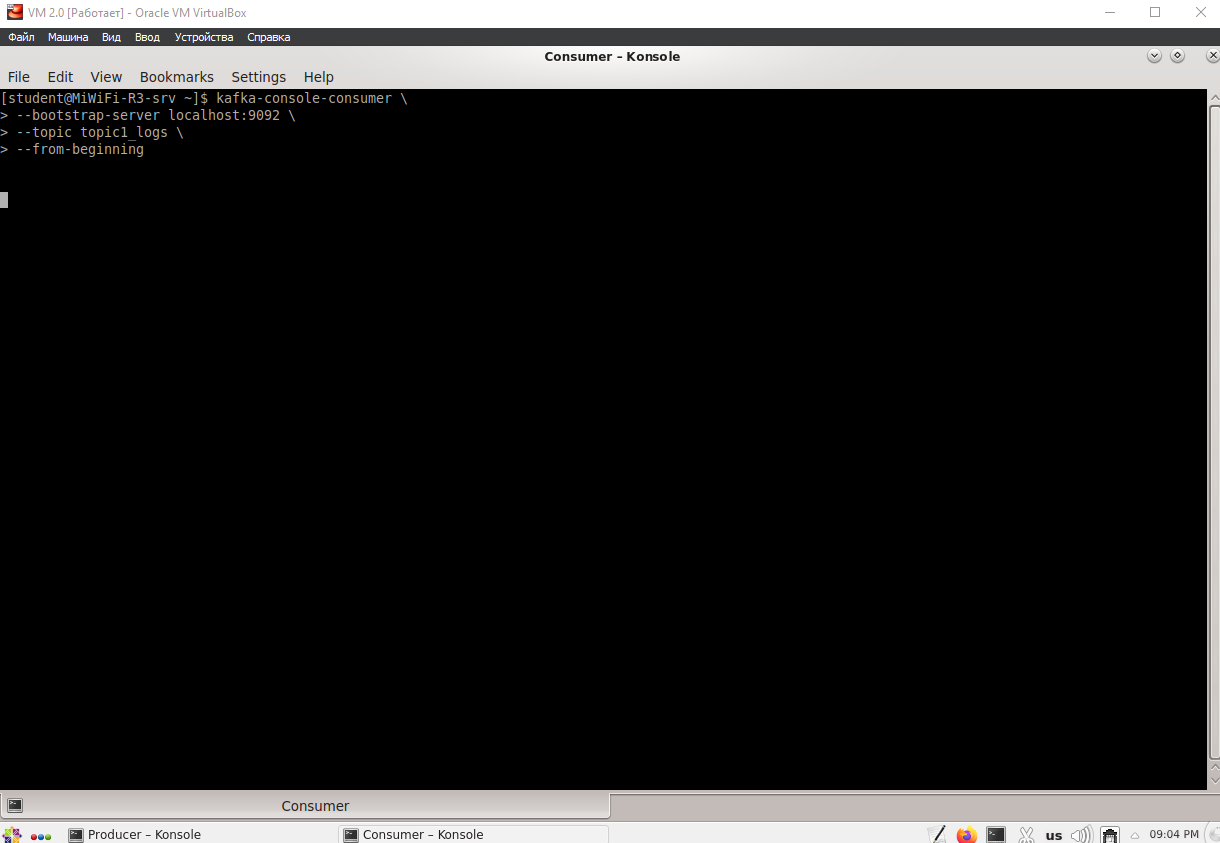


Рисунок 8 — Подготовленный терминал Consumer

Начнем что-нибудь писать в терминале Producer на Рисунке 9.

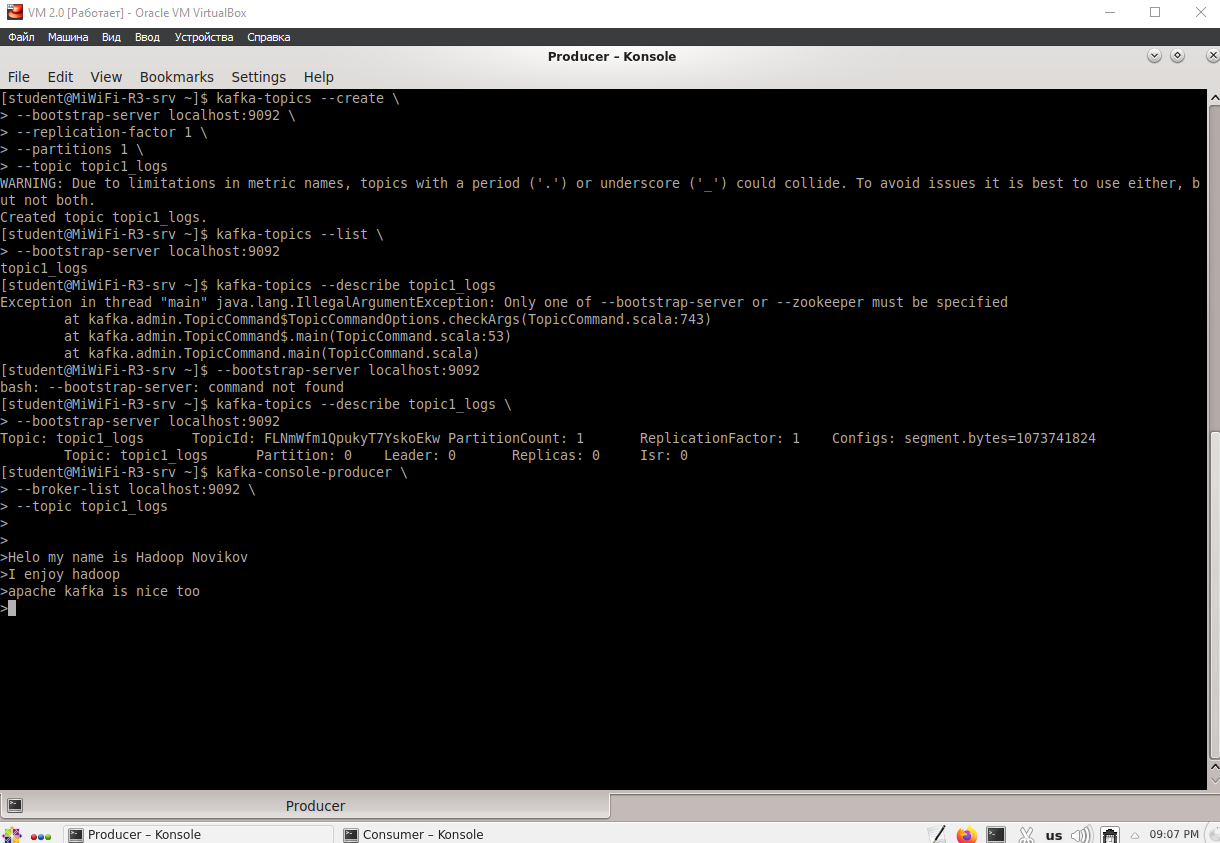


Рисунок 9 — Проверка на передачу текста

Получим текст в терминале Consumer на Рисунке 10.

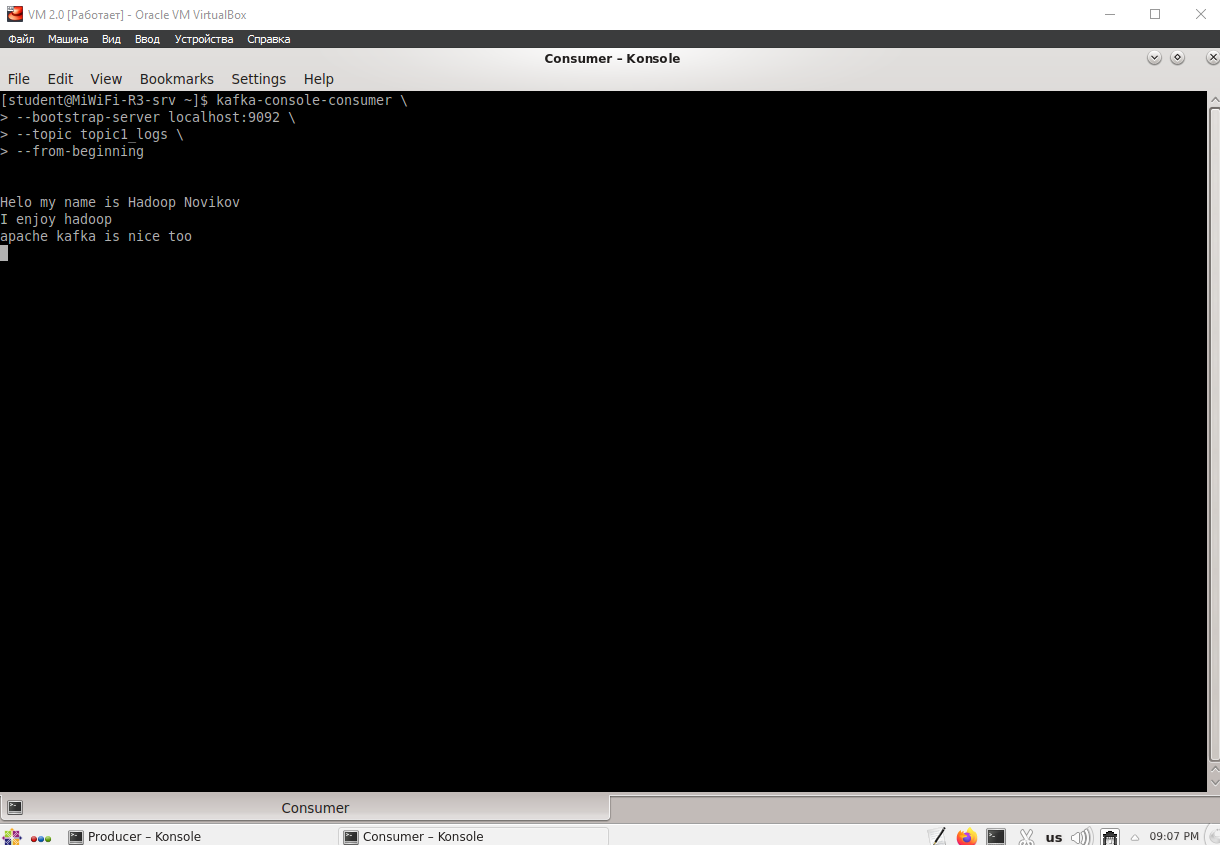


Рисунок 10 — Успешная передача текста

Из терминала Producer остановим производителя нажатием клавиш Ctrl+C на Рисунке 11.

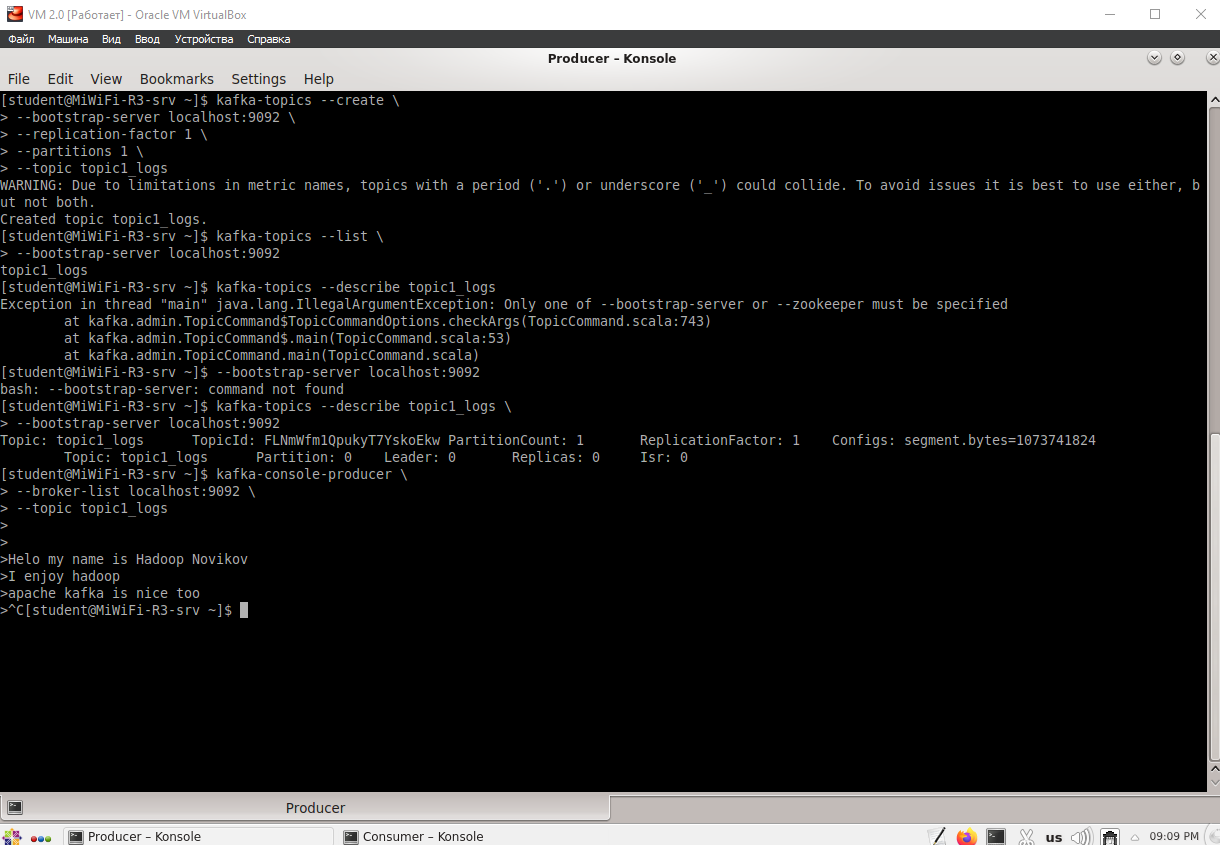


Рисунок 11 — Остановка передачи данных

Отправим содержимое файла Alice-in-Wonderlands.txt в тему topic\_logs на Рисунке 12.

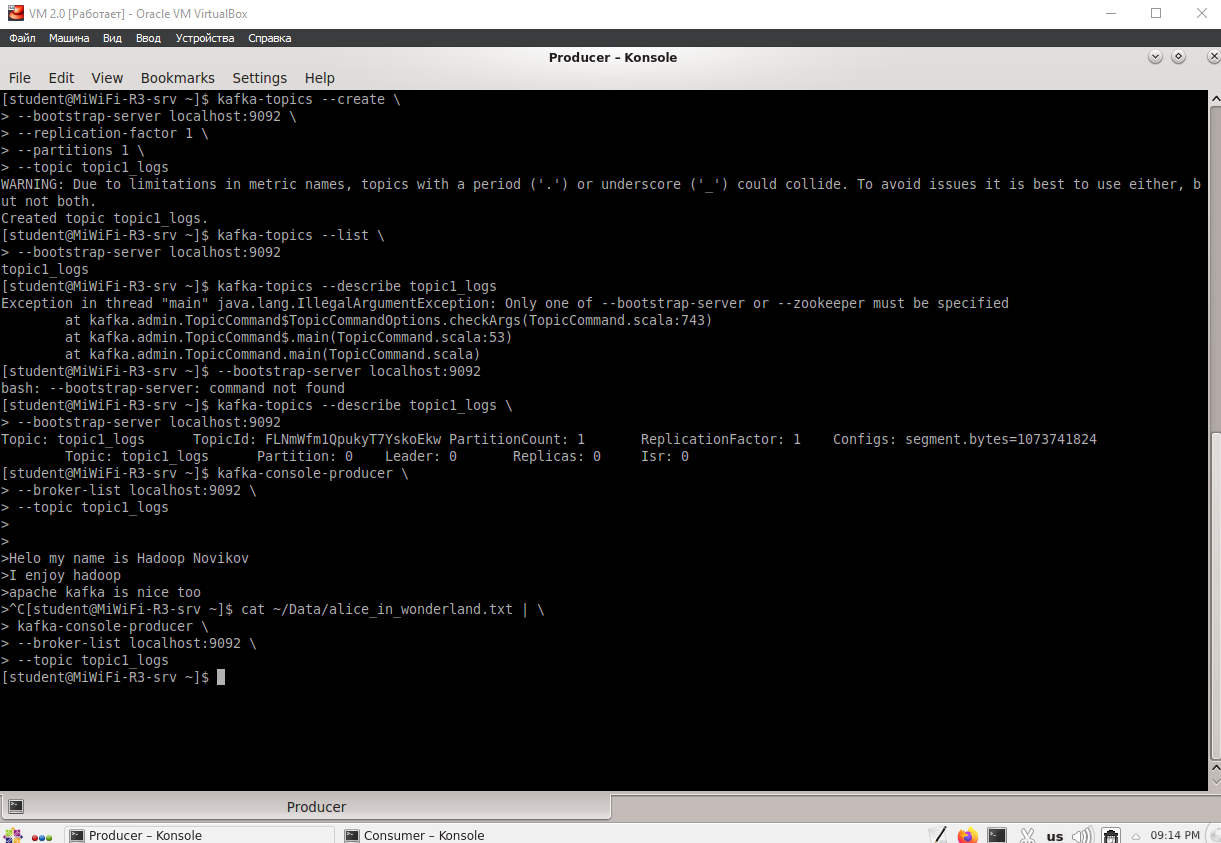


Рисунок 12 — Передача файла Alice-in-Wonderlands.txt

Результат передачи текста представлен в терминале Consumer на Рисунке 13.

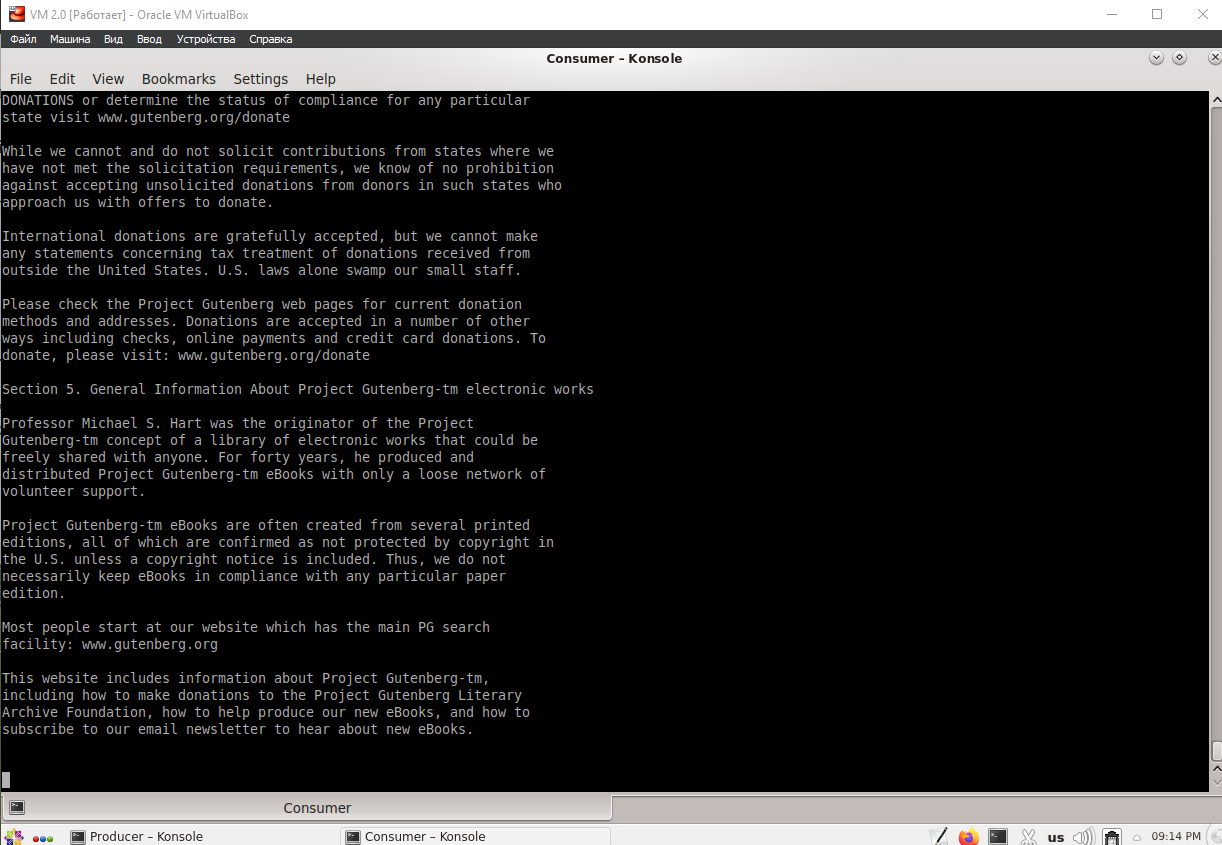


Рисунок 13 — Результат в терминале Consumer

### Отправка сообщений из Flume в Kafka

Как и в прошлый раз, чтобы снизить нагрузку на виртуальную машину, мы остановим HBase и запустим только Apache Kafka на Рисунке 14.

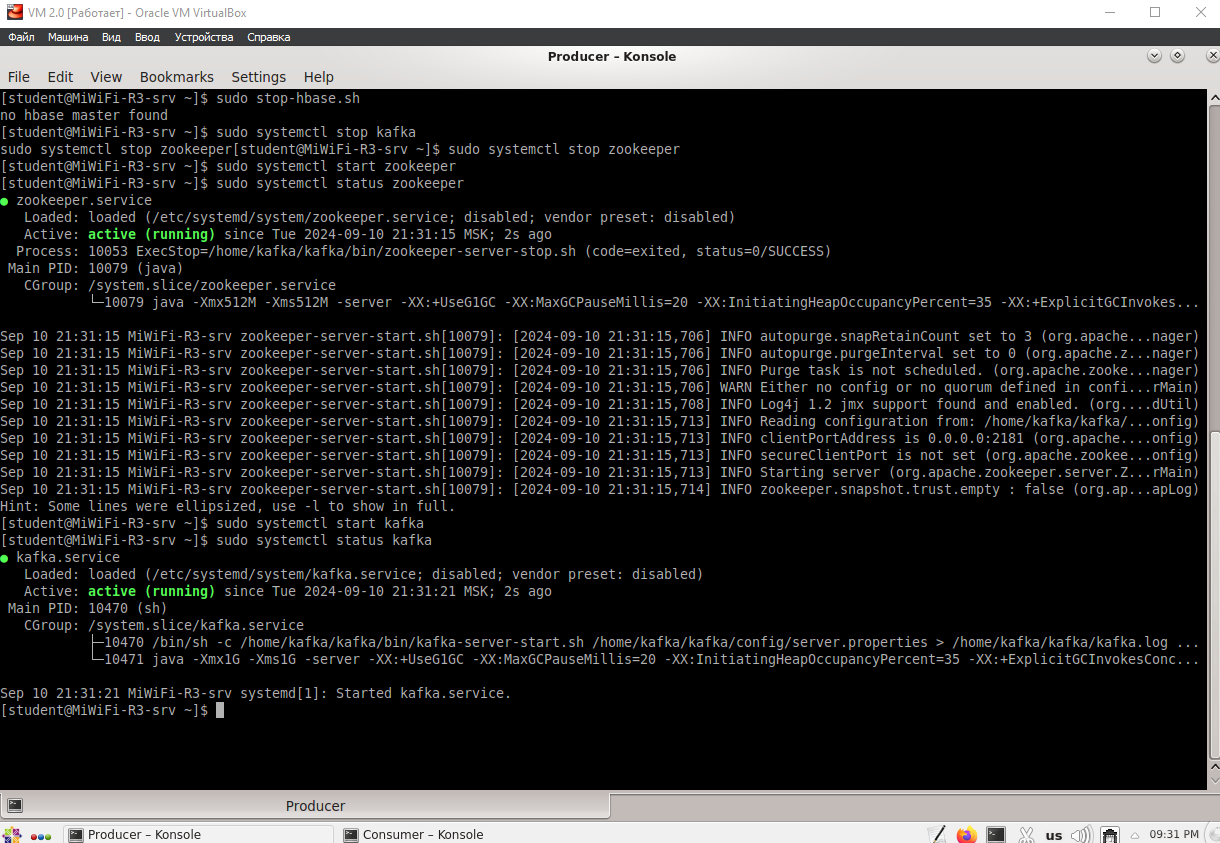


Рисунок 14 — Перезапуск сервисов

Перейдем в папку /home/student/Labs/C3U4 в терминале, а затем создадим новую папку и назовем ее spooldir на Рисунке 15.



Рисунок 15 — Создание новой папки spooldir

Создадим файл конфигурации агента spooldir.conf и заполним его при помощи команды nano на Рисунке 16.

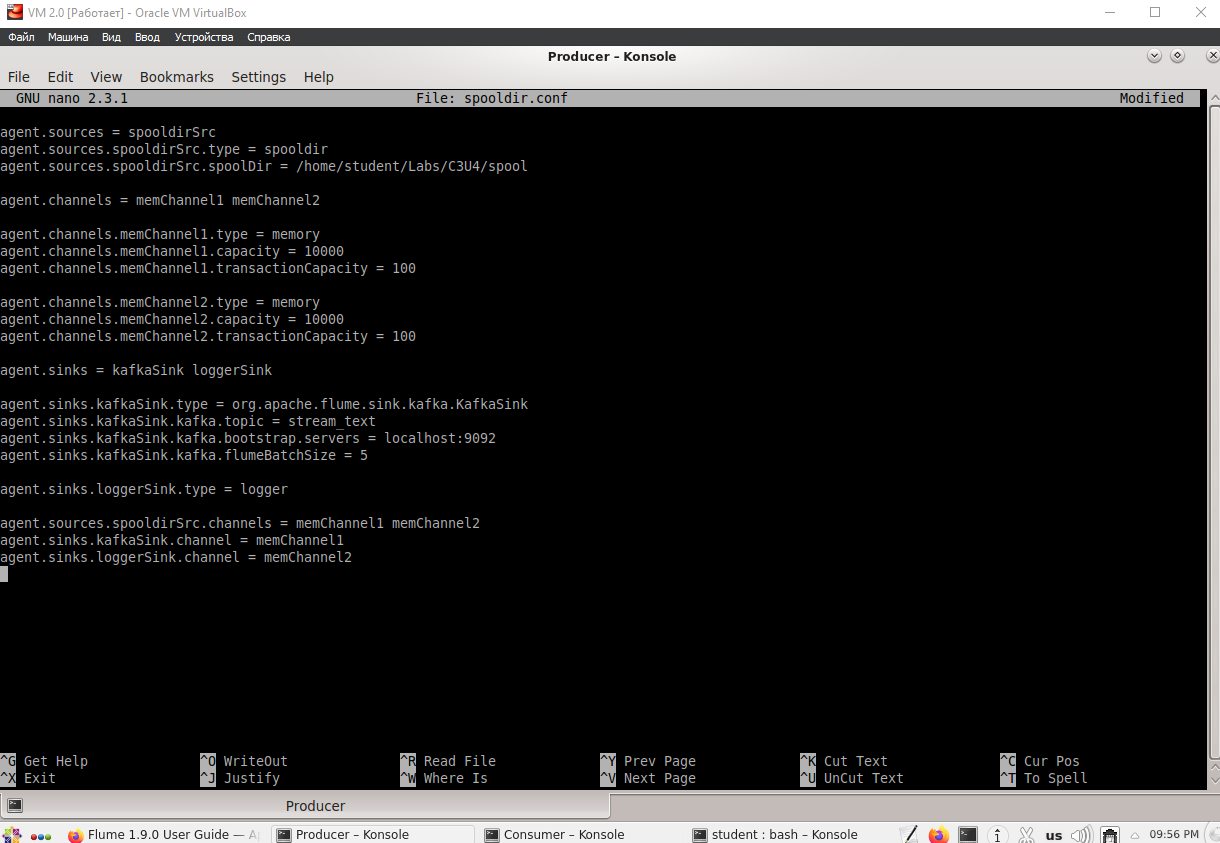


Рисунок 16 — Заполнение файла конфигурации агента

Выполним команду kafka-console-consumer со следующими параметрами: установии значение topic на stream\_text, запустим потребителя так, чтобы он читал сначала (--from-beginning) на Рисунке 17.



Рисунок 17 — Запуск kafka-console-consumer

Перейдем в /home/student/Labs/C3U4. Создаем каталог и назовем его spool. Откроем новый терминал и изменем имя терминала на «Kafka Producer». Запустите поток данных Flume на Рисунке 18.

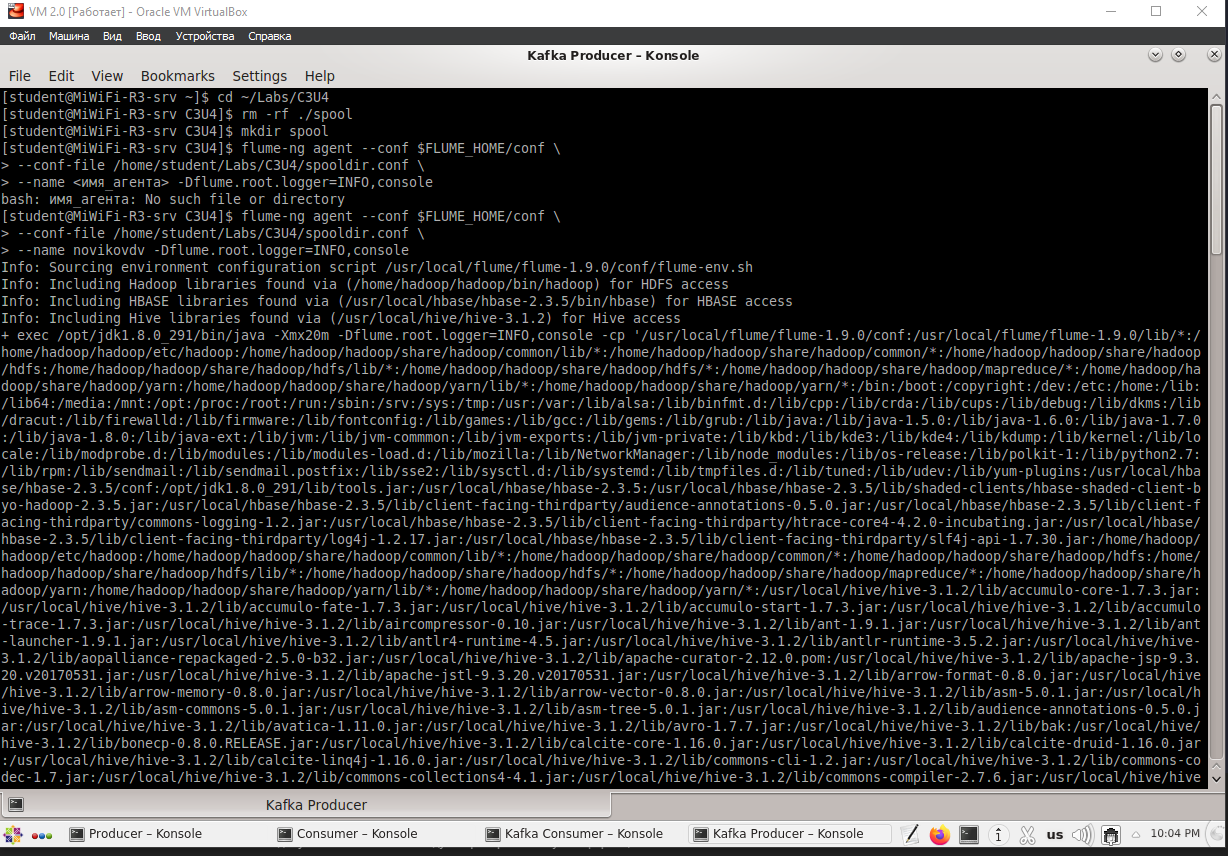


Рисунок 18 — Запуск потока данных Flume

Откроем новый терминал и назовем его «Streaming Source». Перейдем в /home/student/Labs/C3U4. Выполним программу spool\_stream.py (этот файл уже есть в папке C3U4) с помощью интерпретатора python. Следующая команда принимает 5000 символов из alice\_in\_wonderland.txt , создает временный файл для их хранения, а затем перемещает файл в каталог spool. Программа создает один файл каждые 5 секунд на Рисунке 19.

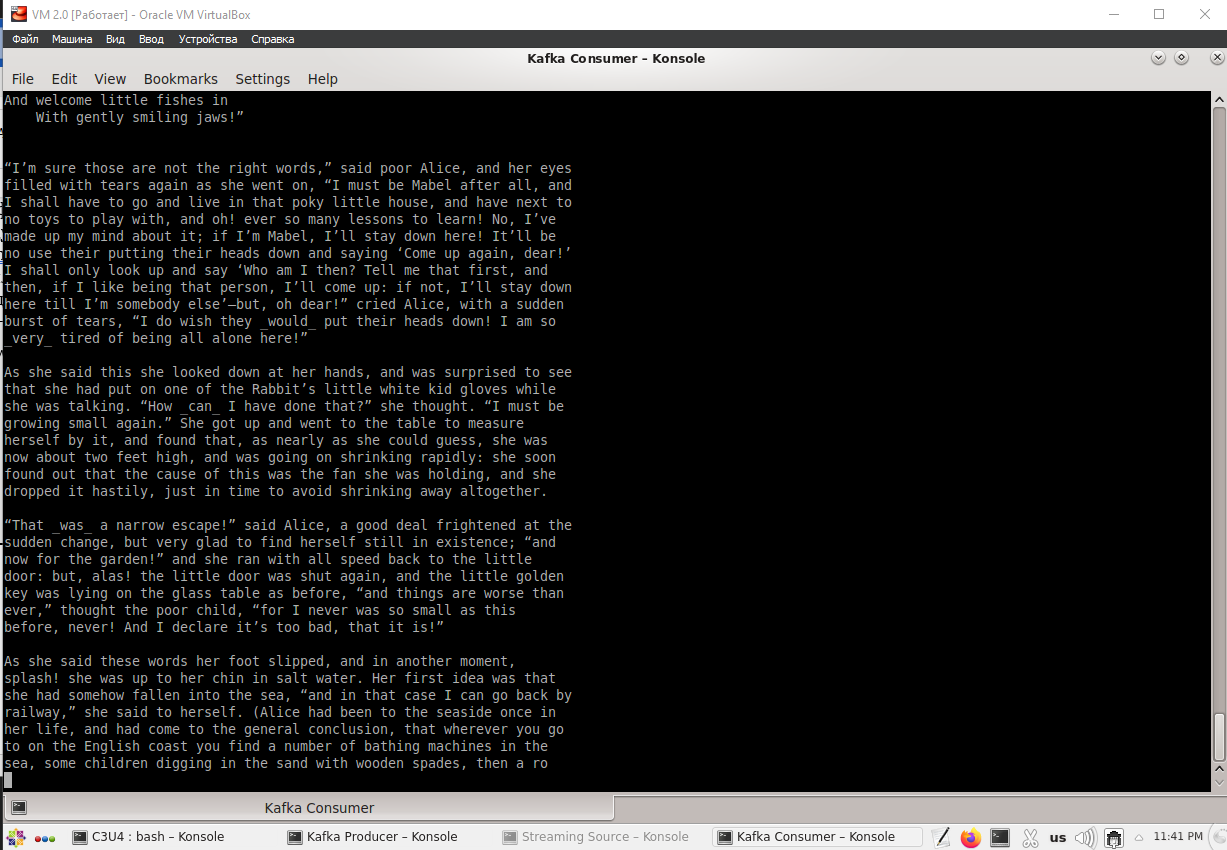


Рисунок 19 — Работа Kafka Consumer

Работа Streaming Source показана на Рисунке 20.

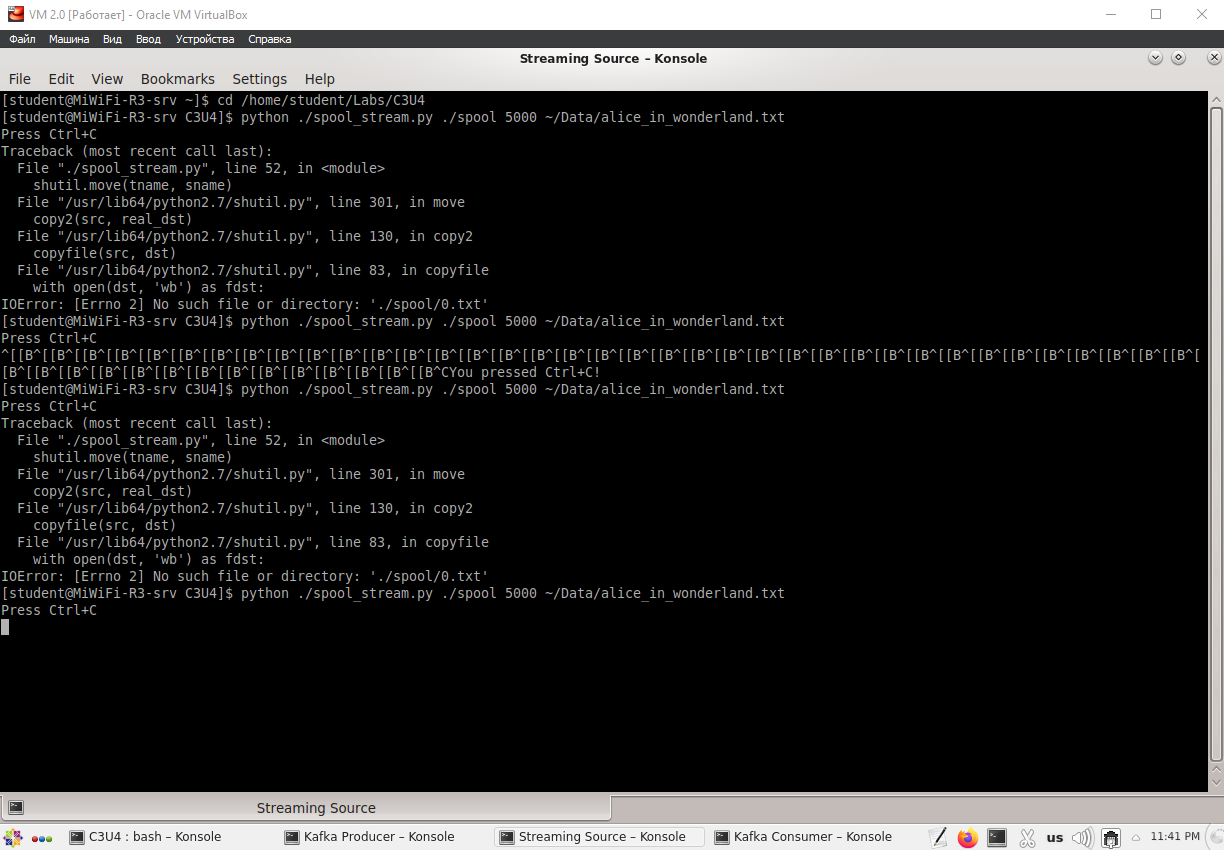


Рисунок 20 — Работа Streaming Source

После остановки посмотрим результат работы Kafka Consumer на Рисунке 21.

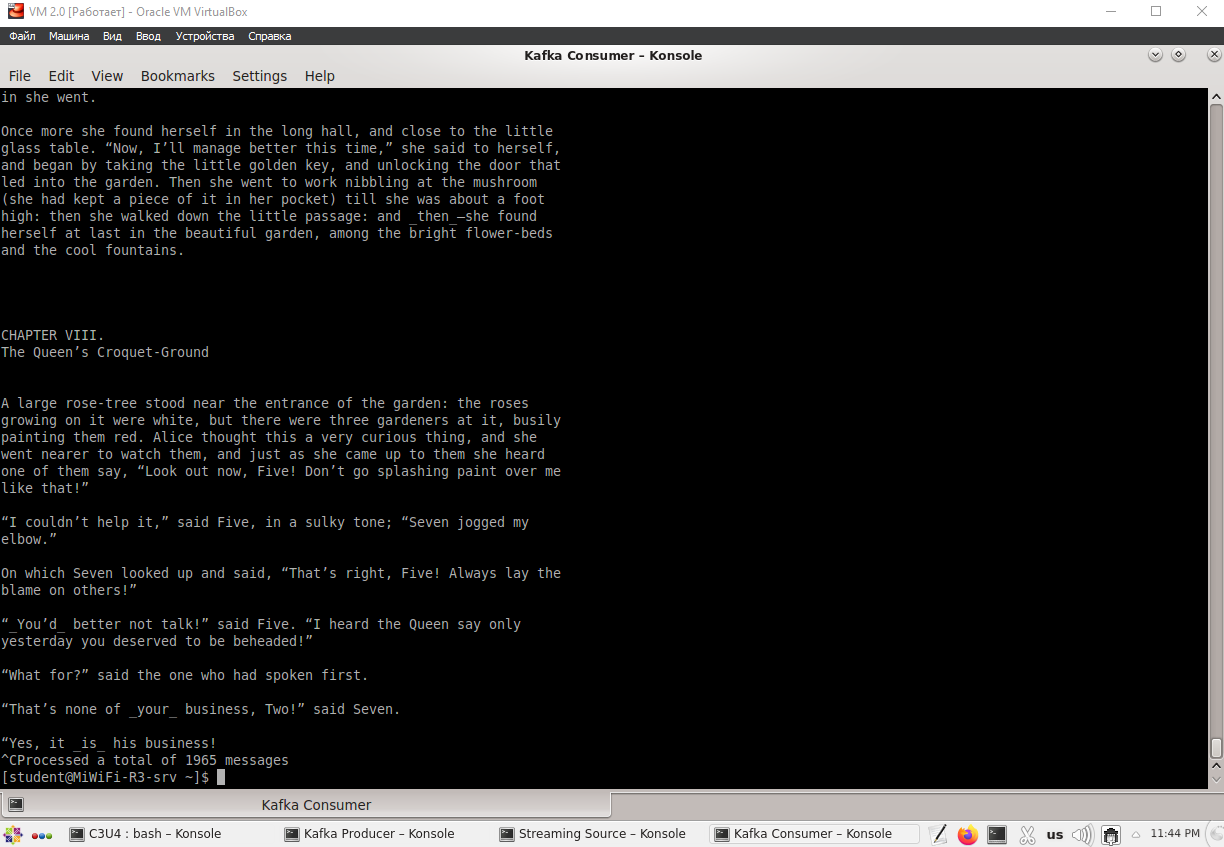


Рисунок 21 — Результат работы Kafka Consumer

Работа Kafka Producer представлена на Рисунке 22.

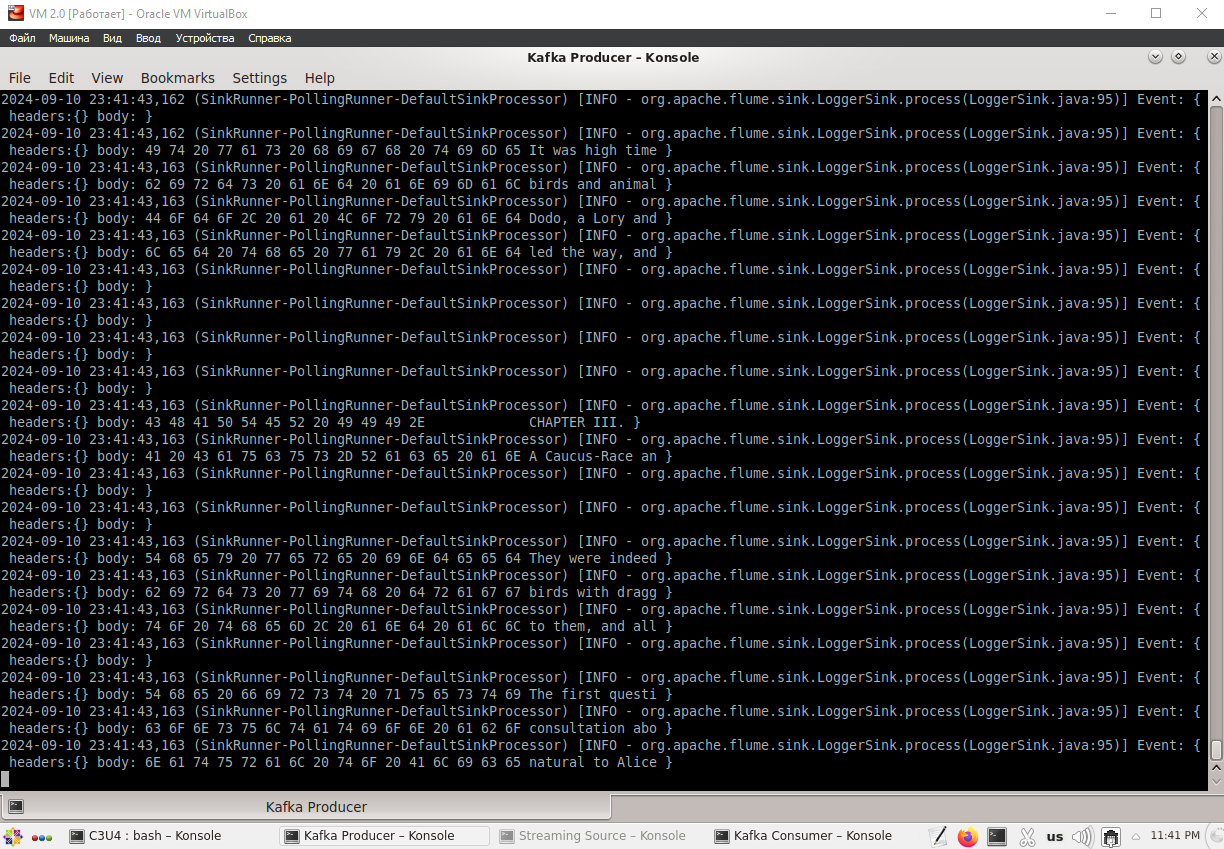


Рисунок 22 — Работа Kafka Producer